УДК 004.934

В.В. Яценко, М.М. Сажок
Міжнародний науково-навчальний центр інформаційних технологій та систем, м. Київ, Україна
val-yatsenko@yandex.ru, mykola@uasoiro.org.ua

Розпізнавання та смислова інтерпретація злитого українського мовлення для усного фразника-перекладача в умовах альтернативних граматик

У статті розглядаються проблеми створення систем смислової інтерпретації мовленнєвого сигналу в межах предметних областей. Описуються базові структури, за якими генеруються еквівалентні речення, що передають певний смисл. Розглядаються три способи побудови породжувальних граматик для розпізнавання злитого українського мовлення: на основі LISP-структур, вільний порядок слідування слів та граматика, заснована на лінгвістичному понятті про фонетичне слово. Запропоновано імовірнісний спосіб формування відповіді смислової інтерпретації. Наводяться результати експериментальних досліджень смислової інтерпретації вимовлених диктором речень, взятих із тематичної області фразника-перекладача.

Вступ

Актуальною є проблема підвищення надійності розпізнавання та розуміння мовленнєвих сигналів. Один із можливих шляхів досягнення цієї мети полягає в економному заданні різноманітних обмежень, зокрема синтаксичних та семантичних, на допустимі послідовності слів у фразах та їх врахуванні при автоматичному розумінні та розпізнаванні мовленнєвих сигналів.

Для чого це потрібно і де використовується? Для різних систем усного діалогу: фразника-перекладача, довідкових систем тощо.

Для слов'янських мов розроблені автоматичні модулі, які забезпечують досить високу надійність розпізнавання, але при цьому вимагають значних обчислювальних ресурсів та фактично прив'язують користувача до одного варіанта висловлення певного смислу. Це пов'язано з тим, що слов'янські мови мають певні відмінності від, наприклад, англійської мови. Основними відмінностями є істотно більша кількість словоформ для кожного слова та відносно вільний порядок слідування слів у фразах. Врахування цих особливостей для слов'янських мов значною мірою ускладнює розв'язування задач розпізнавання та змістовної інтерпретації мовленнєвих сигналів.

Окремою є проблема задання всіх можливих речень мови діалогу, що виражають один і той самий зміст, генерації та пошуку найбільш правдоподібних сигналів та розроблення обмежень на порядок слідування слів, які би враховували лінгвістичні знання.

Для дослідження та специфікації обмежень на допустимі послідовності слів у фразах використовуються LISP-структури [1], [2]. На основі цих структур генерується величезна кількість речень, що мають один і той самий зміст. Втім, існує ряд

обмежень використання цієї технології, пов'язаних як з суб'єктивним чинником при побудові LISP-структур, так і зі збільшенням обчислень, викликаних значним ускладненням графа розпізнавання.

У розділі 1 ми дамо загальну характеристику задач розпізнавання та смислової інтерпретації злитого мовлення, розділ 2 присвячено заданню (специфікації) речень з урахуванням обмежень на допустимі послідовності слів, у розділі 3 пропонується спосіб оцінювання належності послідовності слів до типу речень, у розділі 4 наводяться експериментальні результати.

1. Загальна характеристика задач розпізнавання та інтерпретації злитого мовлення

Розглянемо, в чому полягають і як взаємозв'язані задачі розпізнавання та інтерпретації злитого мовлення [1], [2]. Розпізнавання мови – це процес автоматичної обробки сигналу з метою визначення послідовності слів, які передаються цим сигналом.

Змістовна інтерпретація мови – це процес автоматичної обробки мовленнєвого сигналу з метою виявлення змісту, що передається сигналом, та представлення цього змісту в певній канонічній формі, зручній для подальшого використання. Очевидно, що змістовна інтерпретація мови є більш високим ступенем узагальнення інформації, ніж розпізнавання, оскільки одну і ту саму думку можна виразити різними послідовностями слів. Для отримання кращих результатів розпізнавання та змістовної інтерпретації злитого мовлення ці задачі повинні виконуватися в єдиному взаємопов'язаному процесі. Кінцевою метою цього процесу є зміст повідомлення, який передається послідовністю слів.

Оскільки кожну думку можна висловити різними реченнями в мові діалогу, але при цьому зміст не зміниться, то слід визначити певні обмеження на допустимі послідовності слідування слів у реченнях. Тому при інтерпретації змісту мови різні речення, що передають одну і ту саму думку, повинні відображатися в один і той же результат, тобто відповідь розпізнавання (послідовність слів) не повинна суперечити синтаксису, семантиці та прагматиці предметної області. Зважаючи на це, пропонується розглянути моделі розпізнавання мовленнєвих сигналів, які враховують синтаксис та семантику мови [1], [2].

Задача змістовної інтерпретації злитого мовлення значно складніша за задачу розпізнавання, оскільки для її розв'язання необхідно додатково враховувати апріорну інформацію. Тому перш за все слід навчитися економно задавати всі можливі допустимі речення в мові діалогу. Для вирішення цього питання є декілька шляхів. Один з них – побудова LISP-подібних структур та визначення за їх допомогою обмежень на допустимі послідовності слів.

Цей підхід до розпізнавання та змістовної інтерпретації злитого мовлення може бути реалізовано у вигляді генеративної моделі розуміння (змістовної інтерпретації) злитого мовлення [3].

В рамках генеративної моделі інтерпретація злитого мовлення повинна вписатися в структуру цієї моделі. На рис. 1 можна побачити потужність об'єктів, якими ми оперуємо.

Кожне змістовне висловлювання задамо в канонічній формі, що записана певною семантичною мовою (формальною математичною мовою, що виражає поняття та відношення між ними). Далі за допомогою генератора семантично еквівалентних речень (ГСЕР) задамо перетворення канонічної форми, що не порушує зміст висловлювання. Таким чином ГСЕР породжує всі можливі речення з однаковим змістом,

що визначаються канонічною формою. Далі вводимо перетворення, що породжують всі можливі еталонні сигнали злитого мовлення для кожного речення, згенерованого ГСЕР. Ці еталонні сигнали відрізняються один від одного темпом, що нелінійно змінюється, та інтенсивністю вимовляння.

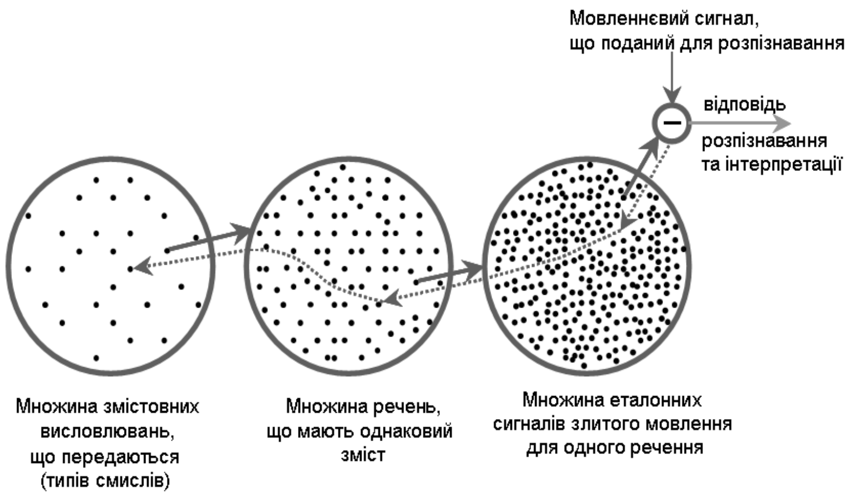


Рисунок 1 – Модель синтезу еталонних сигналів злитого мовлення для змістовної інтерпретації

Автоматичне розуміння (змістовна інтерпретація) пред'явленого мовленнєвого сигналу за допомогою пропонованої генеративної моделі буде полягати в тому, щоб спочатку для сигналу, що аналізується, знайти найбільш правдоподібний еталонний сигнал мови серед всіх сигналів, що породжені генеративною моделлю, а потім визначити канонічну форму того змістовного висловлювання, речення якого відповідає найбільш правдоподібному еталонному сигналу.

Покладається, що задачі розпізнавання та змістовної інтерпретації злитого мовлення повинні розв'язуватися в єдиному взаємопов'язаному процесі, в якому досягається найвища надійність як розпізнавання, так і інтерпретації змісту.

Далі буде розглянуто способи задання всіх можливих речень мови діалогу, що виражають один і той самий зміст, генерації та пошуку найбільш правдоподібних сигналів та розроблення обмежень на допустимі послідовності слів згідно зі структурами, якими можна представити речення.

2. Моделювання обмежень на допустимі послідовності слів

В рамках генеративної моделі для розпізнавання мовленнєвих сигналів запропоновано розглянути певну ієрархію розташування речень. Мається на увазі, що всі мислимі речення мови діалогу розіб'ємо на предметні області (ПО) по типу розмовника для іноземних мов. Кожна предметна область складається із скінченої множини типів змістів (ТЗ). Наприклад, стосовно предметної області щодо відвідання ресторану типи змістів виражаються питаннями про бронювання столика, меню, замовлення тощо. Кожній предметній області відповідає скінчена множина типів змістів. В кожен тип змісту входить множина еквівалентно змістовних типів речень (ТР), які описуються LISP-структурами [1], [4]. Тип речення – це конструкція, що економно задає множину

речень, отриманих з одного речення незалежними допустимою заміною та допустимою перестановкою слів та словосполучень.

Розглянемо приклад ТР для ПО «Повсякденні фрази», що стосується прохання про допомогу у вирішенні проблеми (ТЗ – прохання про допомогу).

$$\begin{bmatrix}Чи\\ *\end{bmatrix}\left(\left(\begin{bmatrix}не\\ *\end{bmatrix}допоможете\right)\left(\begin{bmatrix}Ви\\ *\end{bmatrix}\begin{bmatrix}мені\\ *\end{bmatrix}\right)\begin{pmatrix}вирішити\\ розв`язати\end{pmatrix}(цю\ проблему)\right)$$

В дужках () вказані підсловники, які можна переставляти місцями, а в [] – які не можна переставляти. Переставляти підсловники можна лише всередині старших дужок. Символ * означає порожнє слово.

Неважко переконатися, що наведений тип речення задає 2·4!·2·4·2·1=768 різних речень, допустимих в мові діалогу та таких, що виражають один і той самий зміст прохання про допомогу. Серед цих речень є, наприклад, і такі:

Чи цю проблему не допоможете вирішити Ви мені.
Чи Ви мені цю проблему вирішити не допоможете.

Тобто ми бачимо, що в даний ТР включено багато синтаксично допустимих речень розмовної мови. Але враховуючи вільний порядок слів, серед цих речень будуть утворені речення, які не є типовими для розмовної мови. Тому, щоб відкинути нетипові речення, було запропоновано ввести певні обмеження, які вказують порядок слідування слів.

Всі речення мови діалогу можна задавати за допомогою ТЗ і відповідних їм ТР, використовуючи структуру, наведену у прикладі. За допомогою LISP-структур генерується величезна кількість речень, що мають один і той самий зміст. Оскільки побудова LISP-структур є досить громіздкою, потребує багато ручної роботи, то було розроблено автоматизований специфікатор предметних областей.

Для побудови всіх можливих речень мови усного діалогу будемо використовувати так звану орієнтовану семантичну мережу (ОСМ) [1], [2]. Приклад такої мережі для розглянутого у прикладі типу речення наведено на рис. 2.

Поєднуючи ОСМ типів речень та ОСМ типів змістів, отримаємо ОСМ всієї предметної області. Ця мережа одночасно задає обмежену граматику порядку слідування слів, яку використовуємо при розпізнаванні.

Альтернативною до цієї граматики є граматика вільного порядку слідування слів. Між цими протилежними за суттю граматиками може бути побудовано безліч інших відносно вільних або відносно обмежених граматик. Ми пропонуємо дещо обмежити вільну граматику за рахунок лінгвістичного поняття про фонетичне слово.

Під фонетичним словом розуміємо слово з невіддільними від нього супутніми словами. Наприклад, невіддільними є прийменник від іменника або прикметника, частка «не» спереду дієслова і частка «б» позаду нього.

Щоб формалізувати цю граматику, ми розглядаємо лінгвістичні поняття енклітиків і проклітиків. У групу проклітиків увійшли прийменники (у, в, до, на та інші), сполучники та більшість часток (і, а, де, не тощо), а до енклітиків – частки: ж, же, б, би. Ми також вважаємо за доцільне виділити прийменники з-поміж проклітиків. Остаточно пропонована нами відносно вільна граматика матиме такий вигляд:

(pau <[проклітик] [прийменник | проклітик] нейтральне [енклітик]> pau),

де вміст кутових дужок може повторюватися, а вміст квадратних дужок може бути опущений, *pau* – слово-пауза на початку та в кінці фрази.

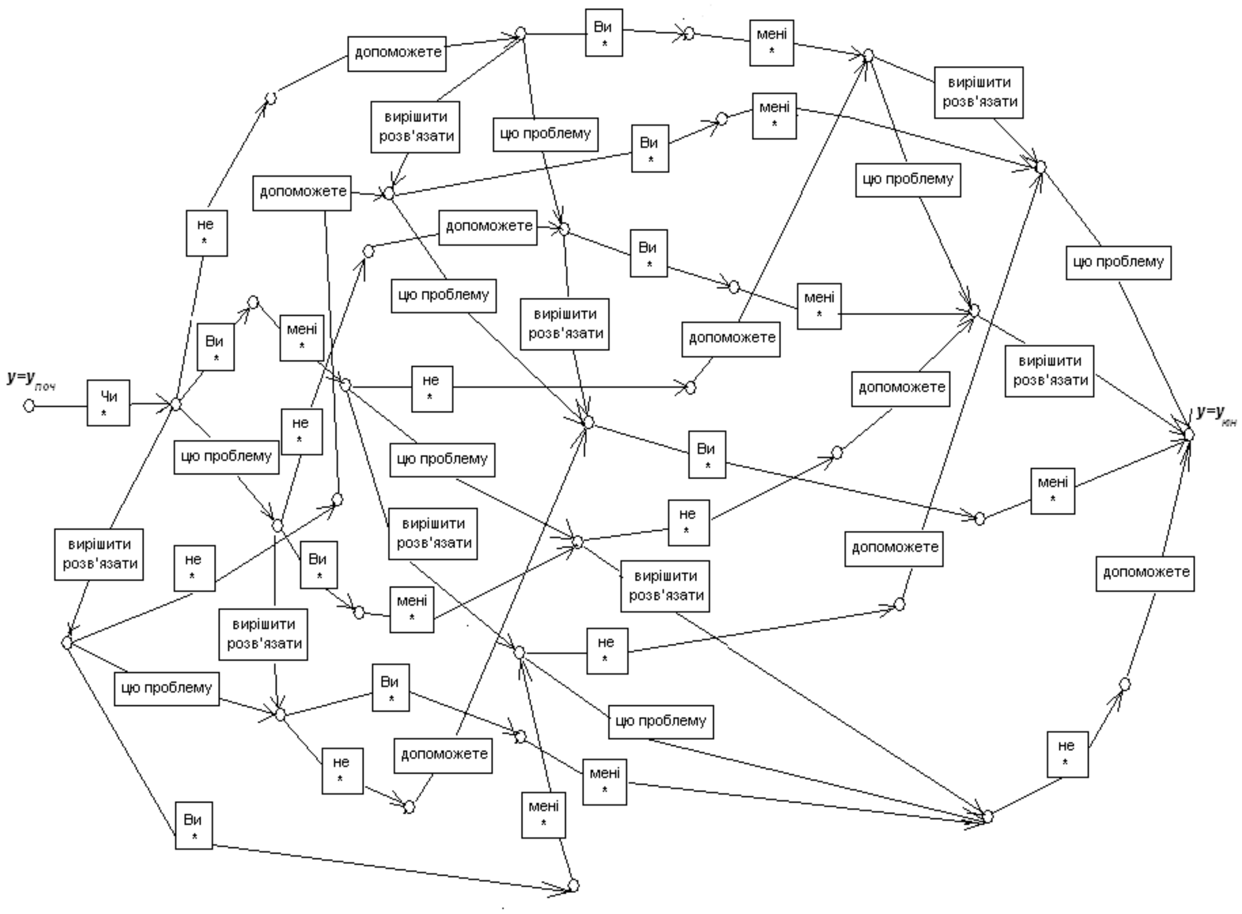


Рисунок 2 – Структура ОСМ для речення, наведеного в прикладі прохання про допомогу у вирішенні проблеми

Втім, за такої граматики прийняття рішення щодо смислу речення не є очевидним. Це будемо розглядати в наступному розділі.

3. Оцінювання належності послідовності слів до типу речень

При розпізнаванні в умовах граматики, що не задає строгих обмежень на послідовності слів, очевидно можуть бути отримані відповіді розпізнавання, що не входять у множину речень, які згенеровані певним типом речень *ST*. Це може бути зумовлено як помилками при розпізнаванні, так і при формуванні типів речень експертом. Крім того, сам користувач може вимовити речення з різного роду відхиленнями або аграматизмами, наприклад повторити деяке слово двічі.

Тому пропонується оцінювати ймовірність типу речення *ST* з ОСМ за умови розпізнаної послідовності слів та оголошувати відповіддю інтерпретації той тип речень *ST**, для якого ця ймовірність є найбільшою:

$$ST^* = \underset{ST}{\operatorname{argmax}} P(ST / w_1, w_2, ..., w_n)\,. \qquad (1)$$

Імовірність у лівій частині (1) може бути переписана за формулою Байєса в такому вигляді:

$$P(ST / w_1, w_2,..., w_n) = \frac{P(ST)}{P(w_1, w_2,..., w_n)} P(w_1, w_2,..., w_n / ST). \quad (2)$$

Якщо припустити рівноімовірність всіх типів речень та рівноімовірність спостережуваних послідовностей слів, то (2) можна переписати в такій еквівалентній формі:

$$P(ST / w_1, w_2,..., w_n) \cong P(w_1, w_2,..., w_n / ST). \quad (3)$$

У свою чергу,

$$\begin{aligned} P(w_1, w_2,..., w_n / ST) = P(w_1 / ST) P(w_2 / ST, w_1) P(w_3 / ST, w_1, w_2) \times .... \\ ... \times P(w_{k-1} / ST, w_1, w_2,..., w_k) \times ... \times P(w_{n-1} / ST, w_1, w_2,..., w_{n-2}) P(w_n / ST, w_1, w_2,..., w_{n-1}) \end{aligned} \quad (4)$$

Розглядаючи послідовність $(w_1, w_2,..., w_n)$ як марківський процес, подаємо кожний із множників у правій частині (4) у вигляді:

$$P(w_k / ST, w_{k-m}, ..., w_{k-1}), \qquad k = 1:n, \quad (5)$$

де $m \geq 0$ – порядок процесу.

Оцінювання кожного з множників виду (5) може виконуватися різними способами в залежності від порядку процесу.

Розглянемо найпростіший випадок, коли $m = 0$. Тоді наша задача зводиться до оцінювання імовірності спостереження кожного з розпізнаваних слів w за умови типу речень ST: $P(w/ST)$. За формулою Байєса цю імовірність представляємо у вигляді:

$$P(w/ST) = \frac{P(w)}{P(ST)} P(ST / w). \quad (6)$$

Як і при аналізі (2), робимо припущення щодо рівноімовірності всіх типів речень, тобто знаменник у цій формулі можемо опустити. За певної умовності рівноімовірними можемо вважати всі слова. Залишається обчислити $P(ST / w)$. Для цього розглянемо $ST(w)$ – множину типів речень, в яких зустрічається слово w. Тоді

$$P(ST/w) = \begin{cases} 0, \text{ якщо } ST(w) = \varnothing, \\ \dfrac{1}{|ST(w)|}, \text{в іншому випадку.} \end{cases} \quad (7)$$

Таким чином, пропонується приймати рішення щодо приналежності розпізнаної послідовності слів певному типу речень на основі (1) – (7). Для апробації цього способу було проведено серію експериментальних досліджень.

4. Експериментальні результати

В якості експериментальних даних було розглянуто англійсько-український розмовник. Розмовник складається з 3800 речень, які назвемо базовими фразами. Ці фрази розділені на 15 предметних областей, кожна з яких має свої типи змістів та типи речень. Для прикладу було розглянуто одну з 15 ПО, а саме «Повсякденні фрази». Ця ПО містить 47 ТЗ та 201 базових речень, в середньому по 5 базових речень на тип

змісту. Щоб задати всю множину речень, яка породжується базовим реченням, кожне базове речення було розмічено відповідно до описаних LISP-структур. Таким чином, для кожного базового речення було побудовано тип речення у вигляді LISP-структури.

Так, наприклад, для типу змісту прохання про допомогу побудовані такі ТР:

*((будь ласка | *) (допоможіть) (мені | *) ((вирішити | розв'язати) (цю проблему)))*
*((будь ласка | *) (допоможіть) (мені | *) (у [цій | *] (справі)))*
((не могли б) (Ви мені) (допомогти) ((вирішити | розв'язати) (цю проблему)))
*[чи | *] (([не | *] [допоможете]) (Ви мені) (вирішити | розв'язати) (цю проблему))*
*((мені | *) (потрібна) ([Ваша | *] (допомога)) (у [цій | *] (справі)))*
*((у цьому питанні) (мені | *) ((буде | *) (потрібна)) ([Ваша | *] (допомога)))*

У наведеному прикладі підсловники містять здебільшого по одному слову, але загалом потужність окремо взятого підсловника може бути більшою в залежності від кількості синонімів.

Було розроблено програмне забезпечення, яке з одного заданого таким чином ТР дає змогу будувати множину всіх речень шляхом відповідних перестановок чи заміни слів та словосполучень. В результаті застосування цієї програми до згаданих вище 201 ТР, було отримано 1045 фраз, не враховуючи змінні параметри. Якщо врахувати змінні, то фраз буде 4337. У словнику нараховується 290 слів.

Для ТР, взятого з наведеного прикладу, – ([чи | *] (([не | *] [допоможете]) (Ви мені) (вирішити | розв'язати) (цю проблему))) – було згенеровано 24 фрази, без урахування змінних, серед них:

$p289 = $w11 $w24 допоможете цю проблему $w19 Ви мені ;
$p295 = $w11 цю проблему $w24 допоможете $w19 Ви мені ;
$p297 = $w11 цю проблему $w19 $w24 допоможете Ви мені ;
$p301 = $w11 $w19 $w24 допоможете цю проблему Ви мені ;
$p304 = $w11 $w19 цю проблему Ви мені $w24 допоможете ;
$p307 = $w11 Ви мені $w24 допоможете цю проблему $w19 ; та інші.

Тут змінними параметрами є $w11=(чи | *); $w19 = (вирішити | розв'язати); $w24 = (не | *). Враховуючи, що кожна змінна може приймати 2 значення, кожна фраза буде мати 8 варіантів. Отже, для даного ТР отримаємо 8·24 = 192 фрази, з урахуванням змінних.

Розпізнавання фраз (речень) проводилося на основі пофонемного розпізнавача за умов обмеженої (на основі LISP-структур), вільної та відносно вільної (на основі фонетичних слів) пословних граматик.

Для експерименту довільним чином було вибрано 100 фраз серед згенерованих 4337, до яких було застосовано алгоритми пофонемного розпізнавання мовленнєвих сигналів в умовах обмеженої та вільної граматик відносно слів [1], [5]. На відміну від попередніх експериментів [4] при оцінюванні акустичних параметрів фонем використовувався корпус кооперативу дикторів. Прийняття рішень щодо смислової інтерпретації здійснювалося на основі (1) – (7). Результати експериментів наводяться в табл. 1.

За умов обмеженої граматики час розпізнавання в 15 раз перевищував реальний час, за умов вільної граматики розпізнавання відбувалося швидше реального часу, а на відносно вільній граматиці – в 1,5 повільніше за реальний час.

Таблиця 1 – Результати розпізнавання та смислової інтерпретації ста речень з предметної області «Повсякденні фрази»

Тип граматики	Результати розпізнавання								
	слів					речень			типів змісту
	%Corr	% Acc	H	D	S	%Corr	H	S	
Обмежена	98,87	98,42	639	5	0	95	95	0	98
Вільна послівна	49,53	45,65	319	193	132	0	0	100	85
Відносно вільна	78,88	74,69	508	14	122	20	20	80	96

Коректність розпізнавання обчислювалася за формулою $\%Correct = \frac{H}{N}100\%$, а надійність: $\%Accuracy = \frac{H-I}{N}100\%$, де:

$H-$ кількість правильно розпізнаних слів/речень,

$D-$ кількість слів/речень, що випали,

$S-$ кількість слів/речень, замінених на інше, порівняно з вимовленим,

$I-$ кількість вставлених слів/речень, яких не мало б бути,

$N-$ загальна кількість вимовлених слів/речень.

На основі цих результатів було розроблено демонстраційне програмне за безпечення для перекладу фрази, вимовленої українською мовою, на англійську мову. При цьому слідування слів в українській фразі може бути будь-яким із допустимих. Фразі, вимовленій українською мовою, ставиться у відповідність англомовний тип змісту або речення, а перше речення цього типу змісту оголошується результатом перекладу.

Висновки

В роботі були розглянуті питання смислової інтерпретації усномовного сигналу з урахуванням специфіки слов'янських мов: відносно вільний порядок слідування слів і їх змінюваність.

Опрацьовано спосіб задання множини допустимих речень, що відповідають одному і тому ж змістові, шляхом побудови типів речень (ТР) засобами LISP-структур. Розроблено програмні засоби побудови граматик слідування слів для розпізнавання злитого мовлення як на основі ТР, так і на основі лінгвістичного поняття про фонетичне слово.

Запропоновано більш гнучкий імовірнісний спосіб формування відповіді смислової інтерпретації, що ґрунтується на припущенні, що спостережувані послідовності слів є марківським процесом.

Експериментальні дослідження показали, що імовірнісний підхід при смисловій інтерпретації показує досить високі результати для обмеженої та відносно вільної граматик слідування слів.

На основі експериментальної моделі розроблено програмну модель усного словника-перекладача для перекладу з української мови на англійську в межах предметної області.

При генеруванні речень за LISP-структурами отримуються в тому числі і речення, які є менш типовими у мовленні. На майбутнє це слід дослідити. Корисним вбачається розроблення алгоритму повністю автоматичної побудови ТР за заданими реченнями.

Одні і ті ж самі фрази з різною інтонацією можуть виражати як питальне речення, так і розповідне. Отже, в подальшій роботі слід дослідити можливість розпізнавання інтонації (просодики) з метою автоматичного розставляння розділових знаків у розпізнаних фразах.

Надалі також планується ставити у відповідність україномовній фразі більш точний англомовний відповідник серед типів речень з типу смислу.

Література

1. Винцюк Т.К. Анализ, распознавание и смысловая интерпретация речевых сигналов. – Киев: Наукова думка, 1987.
2. Винцюк Т.К. Учет синтаксиса языка при распознавании слитной речи. – Киев: Институт кибернетики, 1975.
3. Vintsiuk T., Sazhok M. Multi-Level Multi-Decision Models in ASR // In Proc. Of 10th Int. Conf. «Speech and Computer». – Patras, Greece. – 2005. – P. 69-76.
4. Vintsiuk T., Sazhok M., Yatsenko V. Interpretation of continuous pronunciation for spoken dictionary-interpreter // In Proc. Of 12th Int. Conf. «Speech and Computer». – Moscow, Russia. – 2007. – P. 170-175.
5. Young S.J. et al. HTK Book, version 3.1. – Cambridge University, 2002.

В.В. Яценко, Н.М. Сажок

Распознавание и смысловая интерпретация слитной украинской речи для устного фразника-переводчика в условиях альтернативных грамматик

В статье рассматриваются проблемы создания систем смысловой интерпретации речевого сигнала в рамках предметных областей. Описываются базовые структуры, по которым генерируются эквивалентные предложения, которые передают определенный смысл. Рассматриваются три способа построения порождающих грамматик для распознавания слитной украинской речи: на основе LISP-структур, свободный порядок следования слов и грамматика, основанная на лингвистическом понятии о фонетическом слове. Предложен вероятностный способ формирования ответа смысловой интерпретации. Приводятся результаты экспериментальных исследований смысловой интерпретации произнесённых диктором предложений, взятых из тематической области фразника-переводчика.

Стаття надійшла до редакції 23.07.2008.